



Full length article

Identifying growth morphs from mixtures of size-at-age data

Kyle W. Shertzer^{a,*}, John Fieberg^b, Jennifer C. Potts^a, Michael L. Burton^a^a National Marine Fisheries Service, Southeast Fisheries Science Center, Beaufort, NC, United States^b Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul, MN, United States

ARTICLE INFO

Article history:

Received 30 June 2016

Received in revised form

28 September 2016

Accepted 30 September 2016

Handled by A.E. Punt

Available online 25 October 2016

Keywords:

Bayesian

Cubera snapper

Mixture models

Somatic growth

Unsupervised classification

von Bertalanffy

ABSTRACT

Somatic growth is critical to the biology of individuals and to population dynamics. Variability in size at age can often be attributed to the existence of distinct groups, or growth morphs, that differ in their growth trajectories. We develop a framework for identifying multiple growth morphs from mixture data, with utility for describing somatic growth at the population level as well as for classifying individuals into their most likely groups. For illustration, growth trajectories are modeled using the von Bertalanffy function, but the framework is general enough to accommodate any suitable growth function. After describing the framework, we demonstrate proof of concept using a simulation study, and then apply the proposed method to size-at-age data for Cubera snapper *Lutjanus cyanopterus*. In addition, we compare several Bayesian model selection criteria for inferring the unknown, underlying number of morphs.

Published by Elsevier B.V.

1. Introduction

Understanding somatic growth is fundamental to fishery science. Growth is linked through life-history evolution to reproductive output and natural mortality (Charnov et al., 2013), and these processes are primary drivers of population dynamics. In fish stock assessments, growth models are used to convert between numbers and biomass (in the population and in the catches), interacting with estimation of recruitment, natural mortality, fishing mortality, and selectivity of the fishing or survey gear. Because of their importance to fishery science and consequently resource management, growth models have received much attention in the primary literature (e.g., Maunder et al., 2016).

Numerous models have been proposed to describe fish growth (Quinn and Deriso, 1999; Katsanevakis and Maravelias, 2008). By far, the most widely applied model is attributed to von Bertalanffy (Von Bertalanffy, 1938). The von Bertalanffy model gives rise to an increasing length-at-age, but decelerating growth rate with age, and it fits most populations well, especially when the youngest ages are excluded (Chen et al., 1992; Lester et al., 2004). Alternative models generally lead to a similar growth pattern, though variations exist to account for such influences as life-history stages,

seasonal variation, environmental forcing, and density dependence (Lorenzen, 2016; Matthias et al., 2016).

Somatic growth varies among individuals because of genetic, behavioral, ecological, and environmental diversity. In some cases, sizes of individuals may deviate around a single growth trajectory, as is commonly assumed. In other cases, variability in size or length data may be described best using multiple growth phenotypes, or morphs. Throughout this paper, we use the term “growth morph” to refer to a group of individuals that share an expected growth trajectory.

When fitting growth models to length-at-age data, the typical assumption is that individual fish (observations) originate from the same statistical population. If multiple growth curves are considered, the data are usually first classified into different groups and distinct curves are then fit to each of these subpopulations. For example, if a biological population were expected to have sexually dimorphic growth, separate growth curves would be fit to data from males and females. Then, the two curves could be compared to make inference about dimorphic growth.

Unfortunately, it is not always possible to classify the data *a priori* into distinct subsets. In the preceding example, the sex of some or all individuals may be unknown when fish are gutted at sea (gonads removed), a common occurrence with fishery dependent data. In some cases, assigning labels to groups *a priori* may not be desirable or even possible, but we are nonetheless interested in whether there is evidence for multiple morphs. If so, we

* Corresponding author.

E-mail address: Kyle.Shertzer@noaa.gov (K.W. Shertzer).

may want to estimate the relative abundance of each group in the data set or population, as well as estimate group-specific growth trajectories for use in stock assessment models (Goodyear, 1984; Punt et al., 2002; Taylor and Methot, 2013). In other applications, we may be more interested in classifying individual fish into their respective groups. Whether focused at the population or individual level, mixture modeling can be useful for analyzing data composed of multiple groups, particularly when the group identity of some or all individuals is unknown.

Mixture modeling is well established in the fields of statistics (McLachlan and Peel, 2000; Gelman et al., 2013) and machine learning, where it is called “unsupervised classification” (Hastie et al., 2009). Mixture models have also proven useful in a range of ecological applications – examples include modeling heterogeneity in mark-recapture data (Pledger et al., 2010), temporal variability in animal movement dynamics (Morales et al., 2004), and variation in ungulate life-history parameters (Hamel et al., 2016), as well as decomposing length frequencies into age or stage classes (Sweeney et al., 2015). The latter can be accomplished in a maximum likelihood framework using publically available software such as the R package mixdist (Macdonald and Du, 2012). To our knowledge, mixture models have not previously been applied to somatic growth data collected from fish populations.

The goals of this paper are twofold. First, we show how mixture models can be adapted to account for heterogeneity in group-related growth. Second, we examine ways to identify objectively how many morphs contribute to the population. The method we propose has utility for quantifying growth trajectories of each of the underlying morphs, i.e., the mixture components. It has further utility for classifying individual observations (fish) into their most likely components (e.g., unclassified individuals of a particular length as either male or female). Uncertainty in these classifications can be quantified, allowing for propagation of error to any subsequent analyses. After describing the method, we demonstrate its use on simulated data and through a case study on Cubera snapper *Lutjanus cyanopterus*.

2. Methods

2.1. Mixture model framework

For fitting length-at-age data, we took a mixture modeling approach implemented in a Bayesian framework (Gelman et al., 2013). Here, the observed lengths ($y = y_1, \dots, y_n$) were considered to be generated by a mixture of M components ($m = 1, \dots, M$), the growth morphs. We modeled the expected length ($\bar{L}_{a,m}$) at age (a) of each morph using the von Bertalanffy equation,

$$\bar{L}_{a,m} = L_{\infty,m} (1 - e^{-k_m(a-a_{0,m})}) \quad (1)$$

where $L_{\infty,m}$ is the asymptotic expected length of morph m , k_m is the growth coefficient, and $a_{0,m}$ is the theoretical age at which length is zero. We chose the von Bertalanffy equation because of its widespread popularity in fishery science, but note that the general modeling approach can accommodate any suitable growth function.

We used a latent (unobserved) variable, z_i , assumed to follow a multinomial distribution, to capture the group (i.e., morph) membership of fish i ,

$$z_i \sim \text{multinomial}(\lambda_1, \dots, \lambda_M) \quad (2)$$

Values of λ represent the probabilities of membership in each group and, for Eq. (2) to be a valid probability distribution function, they must satisfy the condition $\sum_{m=1}^M \lambda_m = 1$. We assumed that the λ_m 's follow a Dirichlet distribution. This was accomplished implicitly by specifying a set of M hyperprior random variables,

$$\alpha_m \sim \text{gamma}(1, 1), \text{ followed by the relationship, } \lambda_m = \alpha_m / \sum_{j=1}^M \alpha_j$$

(Kéry and Schaub, 2012). More generally, $\log(\alpha_m)$ could be modeled as a function of explanatory variables in a multinomial regression framework, thus allowing the classifications to be informed by additional covariate data, if available.

We assumed that variability of length-at-age is described by a morph-specific normal distribution,

$$y_i | \theta_m, z_i \sim N(\bar{L}_{a_i,m}, \sigma_m^2) \quad (3)$$

where θ_m is the set of four parameters that fully describe each morph, $\theta_m = (L_{\infty,m}, k_m, a_{0,m}, \sigma_m)$. The assumption of normality is common (Francis, 2016), and we chose it here for the purpose of illustration. However, the approach could easily be adapted to other statistical distributions.

We applied uniform prior distributions on each parameter. For any given application, meaningful bounds on prior distributions will depend on the data set in hand. However, one caveat regarding mixture models is that the z_i might not be identifiable without informative priors or constraints. Indeed, all mixture models are nonidentifiable in the sense that the mixture distribution is unaffected by permutations in the group labels (Gelman et al., 2013). To avoid this ambiguity, we applied the following priors on asymptotic expected length,

$$L_{\infty,1} \sim U(\min, \max), L_{\infty,2} \sim U(0, L_{\infty,1}), \dots, L_{\infty,M} \sim U(0, L_{\infty,M-1}) \quad (4)$$

which is equivalent to the restriction that $L_{\infty,1} > L_{\infty,2} > \dots > L_{\infty,M}$. We define application-specific values of *min* and *max* in the *Simulated data* and *Cubera snapper case study* sections below, as well as specify prior distributions for other parameters of θ_m .

To implement the model, we used JAGS version 4.2.0 (Plummer, 2003), a program for Bayesian analysis utilizing Markov Chain Monte Carlo (MCMC) simulation. It was run in R version 3.2.5 (R Development Core and Team, 2016) via the R package R2jags (Su and Yajima, 2015). We ran three independent Markov chains, each for 500,000 iterations, and each initialized with over-dispersed starting values as judged by viewing the trace plots. Posterior distributions were computed after a burn-in period of 100,000 iterations to avoid any influence of starting values. We thinned the resulting chains by keeping every tenth iteration, because of data storage limits (Link and Eaton, 2012). Convergence was assessed through visual inspection of trace, density, and autocorrelation plots, and by examining the Brooks–Gelman–Rubin statistic for convergence toward 1 (Brooks and Gelman, 1998).

2.2. Classification of observations

During each MCMC iteration, each fish is assigned to a single morph. Across iterations, however, morph assignments may differ. We used the mode of the posterior distribution (i.e., the morph assigned most frequently) to classify each fish.

2.3. Number of morphs and model selection

In some applications, M would be known *a priori*, for example in the case of investigating sexually dimorphic growth where $M = 2$. In other applications, the number of morphs may be unknown. In general, determining the number of components (morphs) in mixture modeling is a difficult challenge (Gelman et al., 2013), and one method for doing so is model selection. For our applications, we fitted the model for multiple values of M , and then compared results graphically and through model selection criteria.

Numerous model selection criteria have been proposed for Bayesian applications, with no general consensus view on which

Download English Version:

<https://daneshyari.com/en/article/4542643>

Download Persian Version:

<https://daneshyari.com/article/4542643>

[Daneshyari.com](https://daneshyari.com)